

STAR SCHEMA - CAMADA GOLD

Modelo Dimensional - Airbnb Rio de Janeiro 2019

Este documento apresenta o Star Schema (Esquema Estrela) da camada Gold do Data Warehouse. O modelo dimensional é composto por uma tabela fato central (FACT_OCORRENCIAS) cercada por tabelas de dimensão (DIM_HOSTS, DIM_LOCATIONS, DIM_PROPERTIES, DIM_REVIEWS), otimizado para consultas analíticas e ferramentas de BI.

CARACTERÍSTICAS DO STAR SCHEMA

- **Tabela Fato Central:**

FACT_OCORRENCIAS contém métricas quantitativas (price, cleaning_fee, etc.) e chaves estrangeiras para dimensões.

- **Tabelas de Dimensão:**

4 dimensões descritivas (Hosts, Locations, Properties, Reviews) contendo atributos qualitativos.

- **Chaves Substitutas (Surrogate Keys):**

Todas as tabelas usam srk_* como chaves primárias, independentes dos IDs originais.

- **Desnormalização Controlada:**

Dimensões podem conter redundância para otimizar performance de consultas.

- **Granularidade:**

Cada registro na tabela fato representa uma ocorrência/listagem de propriedade no Airbnb.

ENTIDADES DO MODELO

- **FACT_OCORRENCIAS** - Tabela Fato
- **DIM_HOSTS** - Dimensão de Anfitriões
- **DIM_LOCATIONS** - Dimensão de Localizações Geográficas
- **DIM_PROPERTIES** - Dimensão de Propriedades/Imóveis
- **DIM_REVIEWS** - Dimensão de Avaliações

RESUMO DE ATRIBUTOS

#	Entidade	Tipo	Qtd	Principais Atributos
1	FACT_OCORRENCIAS	Fato	12	srk_fact_id (PK), 4 FKs, price, cleaning_fee, security_deposit, minimum_night
2	DIM_HOSTS	Dimensão	7	srk_host_id (PK), host_id_original, host_name, host_is_superhost
3	DIM_LOCATIONS	Dimensão	3	srk_location_id (PK), latitude, longitude
4	DIM_PROPERTIES	Dimensão	14	srk_property_id (PK), 2 FKs, property_type, room_type, accommodates, number_of_reviews
5	DIM_REVIEWS	Dimensão	11	srk_review_id (PK), 2 FKs, review_scores_rating, number_of_reviews

Legenda: PK = Primary Key (Chave Primária), FK = Foreign Key (Chave Estrangeira), srk = Surrogate Key

RELACIONAMENTOS (STAR SCHEMA)

- **FACT_OCORRENCIAS (N) —> DIM_HOSTS (1)**
Cada ocorrência pertence a um anfitrião (FK: srk_host_id).
- **FACT_OCORRENCIAS (N) —> DIM_PROPERTIES (1)**
Cada ocorrência refere-se a uma propriedade (FK: srk_property_id).
- **FACT_OCORRENCIAS (N) —> DIM_LOCATIONS (1)**
Cada ocorrência está em uma localização (FK: srk_location_id).
- **FACT_OCORRENCIAS (N) —> DIM_REVIEWS (1)**
Cada ocorrência possui avaliações associadas (FK: srk_review_id).
- **DIM_PROPERTIES (N) —> DIM_HOSTS (1)**
Cada propriedade pertence a um anfitrião (FK: srk_host_id).
- **DIM_PROPERTIES (N) —> DIM_LOCATIONS (1)**
Cada propriedade está em uma localização (FK: srk_location_id).
- **DIM_REVIEWS (N) —> DIM_HOSTS (1)**
Cada avaliação pertence a um anfitrião (FK: srk_host_id).

- **DIM_REVIEWS (N) —> DIM_PROPERTIES (1)**

Cada avaliação refere-se a uma propriedade (FK: srk_property_id).

DICIONÁRIO DE DADOS - STAR SCHEMA

Tabela Fato: FACT_OCORRENCIAS

#	Atributo	Tipo	Nulo	Descrição
1	srk_fact_id	SERIAL	NÃO	Chave substituta do fato (PK)
2	srk_host_id	BIGINT	SIM	FK para DIM_HOSTS
3	srk_property_id	INT	SIM	FK para DIM_PROPERTIES
4	srk_location_id	INT	SIM	FK para DIM_LOCATIONS
5	srk_review_id	INT	SIM	FK para DIM_REVIEWS
6	price	NUMERIC	SIM	Preço da diária
7	security_deposit	NUMERIC	SIM	Depósito de segurança
8	cleaning_fee	NUMERIC	SIM	Taxa de limpeza
9	guests_included	INT	SIM	Número de hóspedes incluídos
10	minimum_nights	INT	SIM	Número mínimo de noites
11	ano	INT	SIM	Ano da ocorrência
12	mes	INT	SIM	Mês da ocorrência (1-12)

Dimensão: DIM_HOSTS

#	Atributo	Tipo	Nulo	Descrição
1	srk_host_id	BIGINT	NÃO	Chave substituta do anfitrião (PK)
2	host_id_original	BIGINT	SIM	ID original do Airbnb
3	host_name	TEXT	SIM	Nome do anfitrião
4	host_response_time	TEXT	SIM	Tempo de resposta
5	host_response_rate	NUMERIC	SIM	Taxa de resposta (%)
6	host_is_superhost	BOOLEAN	SIM	Indicador de Superhost

7	host_listings_count	INT	SIM	Número de propriedades do host
---	---------------------	-----	-----	--------------------------------

Dimensão: DIM_LOCATIONS

#	Atributo	Tipo	Nulo	Descrição
1	srk_location_id	SERIAL	NÃO	Chave substituta da localização (PK)
2	latitude	NUMERIC	NÃO	Coordenada de latitude
3	longitude	NUMERIC	NÃO	Coordenada de longitude

Dimensão: DIM_PROPERTIES

#	Atributo	Tipo	Nulo	Descrição
1	srk_property_id	SERIAL	NÃO	Chave substituta da propriedade (PK)
2	srk_host_id	BIGINT	SIM	FK para DIM_HOSTS
3	srk_location_id	INT	SIM	FK para DIM_LOCATIONS
4	property_type	TEXT	SIM	Tipo de propriedade
5	room_type	TEXT	SIM	Tipo de quarto
6	accommodates	INT	SIM	Capacidade de hóspedes
7	bathrooms	NUMERIC	SIM	Número de banheiros
8	bedrooms	INT	SIM	Número de quartos
9	beds	INT	SIM	Número de camas
10	bed_type	TEXT	SIM	Tipo de cama
11	instant_bookable	BOOLEAN	SIM	Reserva instantânea
12	is_business_travel_ready	BOOLEAN	SIM	Pronto para viagens de negócios
13	cancellation_policy	TEXT	SIM	Política de cancelamento
14	n_amenities	INT	SIM	Número de comodidades

Dimensão: DIM_REVIEWS

#	Atributo	Tipo	Nulo	Descrição
1	srk_review_id	SERIAL	NÃO	Chave substituta da avaliação (PK)
2	srk_host_id	BIGINT	SIM	FK para DIM_HOSTS
3	srk_property_id	INT	SIM	FK para DIM_PROPERTIES
4	number_of_reviews	INT	SIM	Total de avaliações
5	review_scores_rating	NUMERIC	SIM	Nota geral
6	review_scores_accuracy	NUMERIC	SIM	Nota de precisão

7	review_scores_cleanliness	NUMERIC	SIM	Nota de limpeza
8	review_scores_checkin	NUMERIC	SIM	Nota de check-in
9	review_scores_communication	NUMERIC	SIM	Nota de comunicação
10	review_scores_location	NUMERIC	SIM	Nota de localização
11	review_scores_value	NUMERIC	SIM	Nota de custo-benefício